Raptor – a Search Engine Targeting Biodiversity Information

Raptor indexes authoritative biodiversity resources, searching across government organizations at every level, nonprofits, the private sector, educational institutions, and more …

raptor

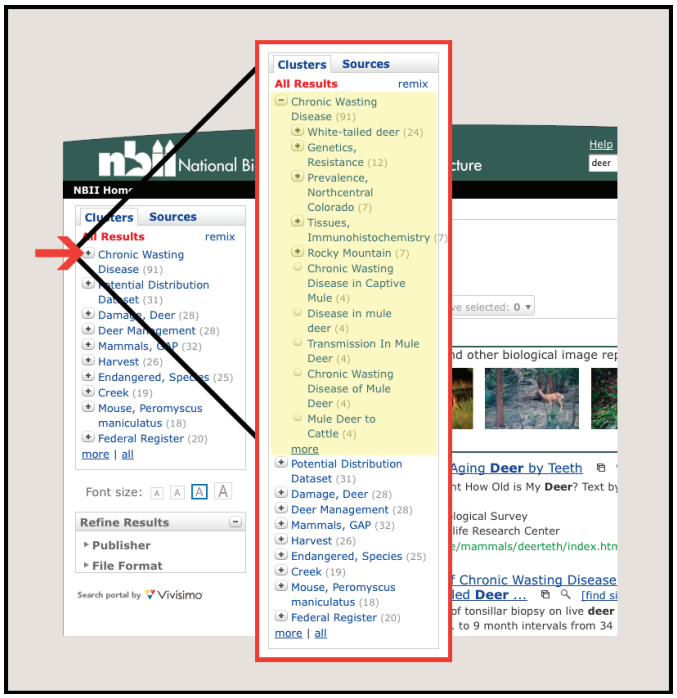
As anyone knows who spends much time looking for information on the Web: searching is good, but finding is better — much better. A new and improved National Biological Information Infrastructure (NBII) search engine has arrived that does both very well.

To see it, just go to the search box in the top right area of the NBII home page www.nbii.gov. There you'll find Raptor, whose name represents what our new search engine is and does for NBII users. A raptor (e.g., an eagle) is a bird of prey that has a mountaintop (comprehensive) view. It scans the terrain, cruises, targets, and retrieves its prey. This Raptor's prey is information, but not just any information.

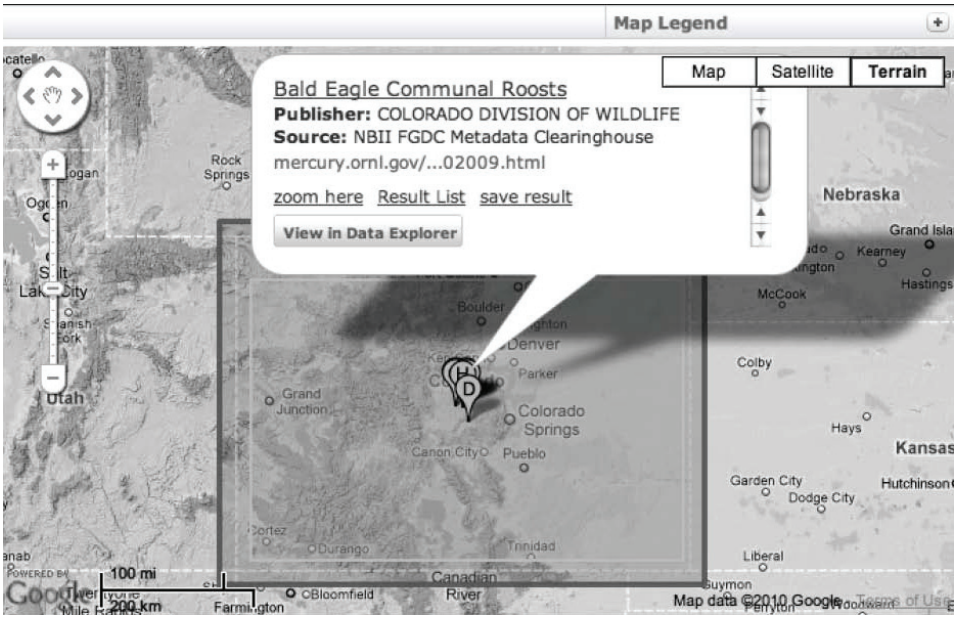
Raptor retrieves precise information for our users, quickly and effortlessly, from more than 40 key repositories of data and information about biodiversity, including all NBII databases and a growing number of documents, Web sites, and databases from authoritative sources outside of the NBII. All of these repositories – which include literature, image galleries, Web sites, databases, and other resources related to biodiversity – have been painstakingly selected from those offered by governments (federal, state, and local), nonprofits, the private sector, educational institutions, and more. Some examples of Raptor's sources include the U.S. Geological Survey Science Centers; U.S. Fish and Wildlife Service collections concerning fisheries, habitat conservation, wetlands, migratory bird management, and threatened and endangered species; the U.S. Forest Service's TreeSearch database of forestry articles and research reports; and the Cornell Ornithology Lab Birds of North America Online Species List. Raptor allows searchers to retrieve information from all of these sources with a single click of the "search" button.

Many Options for Raptor Searches

Users conducting a basic search can enter keywords in the search box and receive results automatically provided by the simplest type of Raptor search. With an advanced search, users can specify which of the repositories to search and which to "turn off." Users can also construct more complex queries, select specific file formats, or define a date range from within the advanced search interface. Many of these choices can also be made from within the results list, making further refinement of search results quick and easy to accomplish.



With Raptor, a search on the term "deer" generates "clusters," a list of related concepts. A user interested in chronic wasting disease, for example, can click on that item in the list and drill down through a new layer of results (highlighted above). This display of "hidden concepts" within a search is a hallmark of the new system.



Raptor's Geospatial search will retrieve and display search results occurring within a user-defined bounding box as "pin drops" on the map. The pin drops are lettered to correspond to the standard, text-based results displayed below the map. Users can click on the pin drops to see basic metadata information and visualize the data associated with it.

Perhaps the most important new feature offered by Raptor is that results arrive in a list and also in "clusters"—thematic categories generated from the texts of searched documents. Users can scan these clusters, weed out irrelevant information, and "drill down" into the areas of interest. For example, a search of the term "deer" will yield clusters about chronic wasting disease, distribution, deer management, and many other themes. Users can then search the themes that serve their needs (see graphic on side 1).

For all searches, the NBII Biocomplexity Thesaurus works behind the scenes to map a user's search terms as well as synonyms to closely related concepts, thus ensuring all potentially relevant sources are made available. For instance, if a user enters "contagious diseases" in the search box, the search will provide matches for that term along with matches for the synonyms "infectious" diseases, "biotic" diseases, and "communicable" diseases. It will also offer the user the ability to explore other concepts that are closely related to the original concept. Related concepts are not included in the initial search result, but are accessible off to the side if the user chooses to click on one or more; in this case, "pathogens" and "viruses." Clicking on one of these will launch a new search based on that term. In short, the new system seeks to understand the context of a user's search without requiring the user to enter keywords for every possible permutation of the topic. To prevent information overload, users have the option to turn off this capability and search only on their designated terms if they prefer.

Geospatial Searching, Too

Raptor users can also find data keyed to a specific geographic region by drawing a "bounding box" (a geographic area delineated by the user to target a search area) on a map. Raptor will then retrieve results occurring within the user-defined region and display it on the map, allowing users to interact visually with the search results.

For instance, suppose you're interested in bald eagles and want to search only in Colorado. Instead of typing in "bald eagle" and "Colorado" into the search box, click on the "Geospatial" option above the generic search box. That brings up the Geospatial search interface, including an interactive map of North America which you can zoom out to search across the entire world or zoom in to select more specific locations. For this example, you would enter "bald eagle" as your search term, then zoom in to Colorado to draw your bounding box (click "draw a bounding box" to draw your box). You have the option to select specific sources to search, or you can leave them all selected, and hit "search" to retrieve your results (see graphic at left).

The purpose is to geographically display search results on a map that are relevant to the search terms, and give users a new way to discover and visualize datasets and information.

Once you see the pin drops on the map, richer data is available with a button called "visualize data." Activating this will open a new application we call the Data Explorer. This powerful tool will visualize map layer data; for instance, geospatially rich NBII products, including the Gap Analysis Program (GAP) landcover map or PAD-US (Protected Areas Database-United States). Searchers can turn these map layers on to visually compare information among them all.

In short, with Raptor Geospatial search, users have the ability to refine their searches geographically, and also to visualize a raw dataset, all in a one-stop shop. If they like what they've found, they can save it as an XML file to their desktop, and perhaps use a geospatial modeling tool to perform new or additional computations.

For More Information

Hugh O'Connor
Senior Digital Librarian
USGS HQ – BRD
12201 Sunrise Valley Drive, MS 302
Reston, VA 20192-0002
Phone: 703-648-4073
E-mail: hoconnor@usgs.gov

Tim Woods
Senior IT Specialist
USGS HQ – BRD
12201 Sunrise Valley Drive, MS 302
Reston, VA 20192-0002
Phone: 703-648-4038
E-mail: twoods@usgs.gov